Connector: A computational approach to study intratumor heterogeneity.

Pernice S(1)[†] , Giordano J(2), Olivero M(2,3), Arigoni M(4), Beccuti M(1), Calogero RA(4), Di Renzo MF(2,3), Sirovich R(5), Cordero F(1)

(1) Department of Computer Science, University of Torino, Torino, Italy
(2) Candiolo Cancer Institute, FPO-IRCCS Candiolo, Torino, Italy
(3) Department of Oncology, University of Torino, Candiolo, Torino, Italy
(4) Department of Molecular Biotechnology and Health Sciences, University of Torino, Torino, Italy
(5) Department of Mathematics G. Peano, University of Torino, Torino, Italy

[†] Email: pernice@di.unito.it

Motivation
The IntraTumor Heterogeneity (ITH) is an essential determinant of tumor progression, diagnosis and treatment [1]. It is the result of the action of the evolutionary forces of mutation and selection. ITH gives rise to cancer cell populations with distinct genotypic and phenotypic characteristics. In particular, an high level of ITH may contribute to the failure of cure, by initiating phenotypic diversity and enabling more aggressive and drug resistant clones. Mathematical oncology is interested in understanding the rules underlying cancer evolution. From the first studies on tumor growth kinetics it was reasonable to consider an exponential increase assumption. Successively it was clear that tumors grow progressively slow as the tumor becomes larger, hence sigmoid functions are more appropriate models. Typical parametric models are the Gompertz's and the logistic's. However, the tumor growth data collected in in vivo experiments are characterized by sparse and irregularly time points. These type of data are not particular suitable to be analyzed by parametric approaches. To cope this issue we consider an approach studied by James and Sugar in [2]. In this paper the authors proposed a model-based procedure for Functional Clustering Model (FCM) appropriate to analyze observations that are sparse, irregularly spaced, or occur at different time points for each subject. We propose a R package called "CONNECTOR" in order to fit and cluster the data with respect to parametric models and also the FCM. It allows us to compare the results and consider the solution given by FCM when the classical models are inadequate to describe the tumor behavior too complicates.
Methods
The CONNECTOR package implements an approach to (i) fit the growth data with respect to the FCM and three parametric models, e.g. Malthus, Gompertz and logistic, (ii) cluster the fitted curves associated with similar behavior.
In the FCM the data are fitted through a combination of cubic polynomial functions. This method is based on projecting the original data onto a cubic natural splines basis. The spline is a piecewise polynomial function, specifically third-order polynomials, which pass through a set of control points, i.e. the input data. The

crucial advantage of FCM is that the splines coefficients are treated as a random variables, in order to cope the issue due to sparsely and irregularly sampled curves. The clustering algorithm embedded in the FCM methods is based on the k-means method. A critical point in this type of clustering is the identification of the number of clusters, k. To determine k we propose three different approaches: (i) the Akaike information criterion (AIC) and (ii) Bayesian information criterion (BIC), to estimate the relative quality of statistical models, and the (iii) the Elbow Method, which represents the ratio of the between-group variance to the total variance into the model.

In the case of the parametric models, the unknown parameters that characterize the growth curves are estimated using the least square method. The fitted curves are then clusterized using the k-means method, where k is determined by the methods reported above.

Finally, we implement a methodology to compare FCM with respect to the parametric models. We propose two measures of cluster similarity. The first one is called Withinness, it measures how much the curves belonging to the same cluster are different to the corresponding mean curve. The second one is called Betweenness and it measures how much the cluster mean curves are different from each other.

Results

We study the phenomenon of ITH in epithelian ovarian cancer (EOC) using patient derived xenografts (PDXs) as preclinical models. These models are based on the transfer of tumor samples directly from the patient into immunodeficient mice. The advantage is that the PDX lines showed functional intra-line heterogeneity, i.e. different growth rate and response to drugs of individual PDXs, suggesting that they might be suitable models to study ITH of EOCs.

The data available derived from a EOC of a single patient propagates in parallel and serial progenies, with a total number of 17 samples distributed on six progenies. These samples were analyzed by CONNECTOR. The fitting was performed by FCM and by three parametric models. The AIC, BIC and Elbow methods are concordant to set k equals to four. Then we obtain one partition for each fitting method, composed by four clusters. The Betweenness and Withinness measures identify the FCM as the best methodology to fit and partition the growth curves. The FMC clusters were also analyzed from the genomic point of view in order to verify the hypothesis of the observation of the clonal dynamics in the PDX lines as a reflection of the clonal evolution of the patients source tumors. We have sequenced, at exome level, individual PDX lines. The analysis of these data reveal a strong correlation between exome make-up and the cancer growth pattern identified in the FCM clusters.

References

1. Swanton C., "Intratumor heterogeneity: evolution through space and time." Cancer Res 2012.

2. Gareth M James and Catherine A Sugar, "Clustering for Sparsely Sampled Functional Data.", Journal of the American Statistical Association 2003.